

DOING SCIENCE ON AUDITORY DISPLAY DESIGN IN THE COCKPIT: MERGING LABORATORY RIGOR AND THE AIRCRAFT COCKPIT ENVIRONMENT

Carol A. Simpson
Psycho-Linguistic Research Associates
485 Summit Springs Road
Woodside, CA 94062, USA
csatplra@earthlink.net

ABSTRACT

This paper will discuss a human-system and application centered approach to the conduct of research on auditory alerting system design, from the perspective of 30 years of human factors research on design principles for aircraft cockpit auditory displays. Too often there is a gap between the results of carefully controlled research conducted in the laboratory and the specific questions raised by auditory display engineers as they design a new auditory alerting system for the cockpit. Absent studies that are representative of the cockpit environment, research findings are often extrapolated to a new design without an understanding by the design engineer of other factors that may influence human auditory perception, signal processing, and cognitive interpretation. Alternatively, again in the absence of research findings applicable to the cockpit environment, the design engineer may present some alternative auditory signal designs informally to one or two project pilots, obtain their preferences and suggestions, and design the system to satisfy this small, unrepresentative sample of the user population. Even some of the current standards for auditory display design contain guidance that does not adequately take the cockpit environment into consideration. Examples will be presented with lessons learned and with recommendations for methods of incorporating the rigor of laboratory experimental design into applied research conducted in the aircraft cockpit environment, simulated and real.

[Keywords: Methodology]

1. INTRODUCTION

Despite our best efforts to conduct rigorous research that will lead to good design of auditory displays, displays that are easily perceived, understood, and interpreted by human users, we continue to see new hardware and software that violates the results of earlier research, research that we would expect to have been considered by the design engineers of a given system before and during the design and development of a new auditory display. There are several links from the laboratory to the completion of production systems that are too easily overlooked or not completed. While we can fault the engineers for not paying attention to our work, much of the fault is our own. Our studies fit well into our own basic research plans, each ideally building upon our work and that of our colleagues. But, in our efforts to carefully control experimental variables and possibly confounding variables, we often conduct experiments that are so far removed from a representative application that it is too

difficult for an engineer to apply our results to a particular system design.

We have design guidelines, design standards, and design specifications which are developed and written by consensus among a selected group of experts in the field. But these documents usually lag several years behind the forefront of our research and are prone to being technology-dated.

In addition to our laboratory research in which a few variables are tightly controlled, we need to follow up with rigorously controlled research that at least simulates, better yet is conducted in the application environment with much of the experimental design "noise" that this entails. We need our laboratory research to help us understand the details of human perception, auditory processing, and cognition. We also need the laboratory to help us determine which of the many variables we are manipulating actually have large effects on human performance, effects that are both statistically significant and of a large magnitude. I will call these the strong variables.

We, as a research community of many, then need to conduct application-oriented research that tests the effects of these strong variables to determine if their effects are so strong as to make a difference for humans who are trying to accomplish a task with an auditory display for information input. It is the strong differences that pass the test in both the laboratory and in application-oriented research that should form the basis for engineering design guidelines and standards for auditory displays. The need to close this research gap is likely valid across the board for applications of auditory display research to actual systems that will be used by human operators. This paper focuses on auditory display design for aircraft cockpits as one of many such applications.

2. TWO EXAMPLES OF THE RESEARCH GAP

The author's research has focused primarily on design principles and requirements for auditory displays that are to be used in aircraft cockpits, both civilian and military, both fixed-wing and helicopters; however, auditory alerting signals have long been used for medical monitoring equipment, and now that audio displays are making their way into land vehicles (cars and trucks), it can be presumed that these issues are present in these industries also.

2.1. From Laboratory to End User

There is now a large literature on acoustic features of soundimages or auditory icons and the corresponding meaningful

associations of these features for human listeners. Indeed this very conference is dedicated to immersing users of auditory displays in "Organized Sound".

Patterson's work in the early 1980s exemplifies the careful laboratory-to-development approach to auditory icons for a particular purpose – attention-getting sounds (attensons) for alerting air crew [1]. After testing the effects on perceived urgency of a range of parameters, including pitch, pulse rate, dissonance, Patterson used the test results to develop a set of attentons with four levels of urgency. Tests conducted at the Royal Aircraft Establishment (RAE) Farnborough confirmed the perceived urgency for the Patterson attenton set [2]. The assignment of individual attentons to specific cockpit alerts, such as an engine failure or a low altitude condition, was arbitrary except for mapping perceived urgency level of an attenton into the operational urgency level of the assigned alerting condition. In a full mission tactical helicopter simulation study, Simpson, Williams, Rood, James, and Gardner in 1991 (reported in Simpson and Gardner 1998) [3] compared two alternative design philosophies to cockpit auditory alerting systems: spoken alerts with a preceding attenton, and those same spoken alerts with no preceding attenton. Attensons were assigned to individual spoken alerts or classes of spoken alerts based on the urgency mapping described above. For all conditions, the auditory alerts were presented with head-related auditory localization for spoken alerts pertaining to enemy threats. No localization cues were added to the air vehicle alerts such as engine failure. Pilots trained to a criterion of perfect recognition of each attenton-alert meaning pair prior to flying the simulated missions. After completion of several days of flying simulated, high workload combat missions with the attentons and spoken alerts, the pilots were tested for recognition of the assigned meaning of each attenton. Only one of the twelve attentons was consistently correctly recognized by the pilots, the attenton that had been assigned to indicate that an enemy threat was detected by aircraft sensors. The pilots reported that this attenton was easy to remember because it actually sounded like a particular type of threat.

Based on these pilots' reports, Simpson and Gardner in 1992 [4] developed a method for finding such sound-meaning correspondences by tapping pilots' existing associations with environmental, man made, and computer-generated sounds. The cockpit alerts from an Army helicopter that was then under development were reviewed and a subset of these selected as alerts for which alerting sounds were sought. One hundred recorded sounds from sound effects collections were screened by subject matter experts, in this case helicopter pilots, to select sounds that might "mean" one or more of the cockpit alerts. The resulting screened set of thirty-six sounds was then tested with Army helicopter pilots using a multiple choice response set design. Pilots received no training or exposure to the sounds and potential meanings prior to testing. Two measures were collected for each sound. Each sound was played individually, and pilots selected the best sound-meaning match for that sound from a list of possible meanings. One of the permitted responses was "no match". After a pilot had selected his best sound-meaning match for a given sound, he then rated the strength of the sound-meaning association, called "degree of soundimagery" on a scale from 0 to 10, with a rating of 0 indicating there was no association and a rating of 10 indicating that there was an excellent association of sound and meaning.

Two criteria were established for assessing the strength of each sound-meaning pair. For each sound-meaning pair, the number of pilots who selected that pair was called the "number of votes". A statistical confidence level of $p < 0.01$ was used as the criterion for the number of votes for a given sound-meaning pair. The rating from 0 to 10 was called the "soundimagery rating" for that pair, and the criterion for the soundimagery rating for that sound-meaning pair was that the pilots' mean rating for that pair be greater than the mean rating across all pilots for all sounds. Sounds that met at least one of these criteria were tested in a second study using the same methodology, but in which both criteria had to be met in order for a sound-meaning pair to be selected for use in the auditory alerting system of this helicopter. The seven sounds that resulted from this pair of studies were called "soundimages" by the authors, after the first known use of that term by Archie Sherbert of Boeing Helicopter Company in Philadelphia at the SAE Second Aerospace Behavioral Engineering Technology Conference, Longbeach, CA, 3-6 October, 1983.

In a subsequent study for a different Army helicopter, Simpson and Gardner in 1998 [3] included one of the seven soundimages from the 1992 study plus four of the sounds that had met the votes criterion but not the soundimagery rating criterion, for the meanings "new digital message received" and "master advisory" (advisory attenton). Five of the existing auditory alerting sounds for this helicopter were also included in the test set for comparison and to determine if any confusability among these and the new sounds existed. Using the same methodology as in the earlier study, it was found that the soundimage for "new digital message received" again passed the selection criteria. None of the other four candidate soundimages met the criteria for selection. An interesting added finding was that only two of the existing sounds developed by industry for this helicopter exhibited intrinsic meanings that were the same as their assigned meanings, as indicated by the failure of the other three existing sounds to meet the votes criterion for their currently assigned meanings. One of the current sounds actually met the votes criterion for a different meaning: "master advisory" instead of its assigned meaning in the current cockpit of "master caution". The experimental soundimage for "new digital message received" had a better sound-to-meaning correspondence for helicopter pilots than did the current actual aircraft audio alerts, despite these pilots having been previously exposed to the current audio alerts while flying their helicopter and having never been exposed to the experimental soundimages prior to the testing.

The methodology described here has the advantage of testing sound-meaning pairs for intrinsic soundimagery without prior training. It is, however, weak in its ability to deliver results for any particular meaning for which a soundimage is sought. There is a research gap between the studies that systematically measure and test effects of specific acoustic features on listeners' perceived meanings for sounds and the type of work described here that tests candidate sounds with representative users without any a priori or hypothesized sound-meaning assignments. Used in conjunction with one another, these two approaches can help fill the gap between basic research in the laboratory and applied research for a particular application and user population. Acoustic parameters that are found in the laboratory to reliably convey specific features of auditory icons can be manipulated for sounds that do not quite pass the criteria for soundimages. These modified sounds can then be tested using the soundimage methodology described here. In a complementary fashion, our

soundimage "discovery" methodology can be used to find environmental sounds that have intrinsic meanings for a particular user population; these good soundimages can then be used in laboratory studies that manipulate individual acoustic features to assess the effectiveness of these features. For example, Stevens, Brennan, and Parker reported in 2004 a statistically significant main effect for listeners' accuracy in identifying the identity of each of four auditory icons which were manipulated for size, distance, and direction of motion [5]. Controlling for the effect of identity by using soundimages that meet the criteria described here could in future studies eliminate the interaction of identity with manipulated variables and permit greater experimental control over the variables of interest.

2.2. Ad Hoc Selection of alerting sounds

The literature is replete with examples from earlier times of arbitrary assignment of sounds to cockpit alerting meanings. Unfortunately this practice continues in the avionics engineering community. Recently the author discovered that engineers developing a new cockpit alerting system had by chance selected the "new digital message received" soundimage for use in their system to mean "missile fired at ownship". They made their selection without considering the context of other alerting sounds in this cockpit and without consulting the already existing meanings for these sounds. A small group of test pilots were exposed for the first time to this inappropriate use of this particular soundimage. During the short, two-month period during which these pilots flew tests with the developmental version of this alerting system, they came to associate the "new digital message received" soundimage with "missile". However, a few months post flight test, the project test pilot could not remember the "missile" sound's characteristics. In contrast, over the course of several years of studies with helicopter pilots, the author and her colleagues consistently have found that this soundimage means "new digital message received" for more than the 100 pilots tested to date. Fortunately in this case the author was able to convince the project engineers to remove this inappropriate missile sound from the alerting system. This example illustrates the need for a repository of alerting sounds that are already in use or that have been found to have high soundimagery for a certain alerting meaning. Such a repository could be accessed by auditory system design engineers when selecting a new soundimage or auditory icon for their application.

3. EXAMPLES OF METHODOLOGIES

The soundimage development and test procedure described above is one of the methodologies that the author and her colleagues have used effectively with end users. This soundimage "discovery" methodology treats the end user as a black box, the perceptual and cognitive workings of which we are unaware. But the soundimages that it does produce have been found to be extremely robust across several samples of the user population and across several years of testing in successive studies. In addition to yielding sounds with high soundimagery, or as Stevens et al would say, good identity recognition, this methodology, with the combined votes and soundimagery rating criteria, can serve as a method of validation with end users of a set of soundimages or auditory icons that are proposed for an end user application.

Another type of gap between laboratory research and the application of these results to an application design occurs when results from a carefully controlled and relatively benign laboratory environment are applied to a high workload, high stress application such as a military aircraft in either actual or simulated combat. For example, an auditory icon with a 6 second duration may have produced excellent recognition by listeners in a relaxed laboratory setting but may be annoying and disruptive to a fighter pilot in the midst of an engagement with an enemy. Yet it is expensive and increasingly difficult to obtain funding for full mission simulations or even part task or partial mission, realistic simulation studies. High workload, highly engaging personal computer combat simulations and games offer a low cost method for loading pilots during the testing of alternative designs for auditory alerting systems. The pilots' immersion in and dedication to the task can easily be increased by making the test "engagements" be competitive among the pilots by posting the scores for all test participants to see [6]. A verification of the successful simulation of a high stress environment can be obtained from pilots' post "mission" ratings of workload and of the realism of the simulation.

4. SUMMARY

This paper has discussed some of the sources of and a few solutions for a gap that exists between the results of carefully controlled laboratory research on soundimage and auditory icon design and the end application of these results to the design of auditory alerting systems. Some of the methodologies that can be used to bridge this gap are described, and a recommendation is made for the establishment of a repository of soundimages and auditory icons with their meanings for use by design engineers and the research community.

5. REFERENCES

- [1] R. Patterson, *Guidelines for auditory warning systems on civil aircraft*, CAA Paper 82017, Civil Aviation Authority, London, England, 1982.
- [2] J. Chillery and J. Collister, *Perceived urgency of warning signals determined using a force choice pair comparison technique*, Royal Aerospace Establishment Technical Memorandum MM 12, Ministry of Defense, Farnborough, England, 1988.
- [3] C. Simpson and D. Gardner, *Soundimage for IDM message receipt: Evaluation of candidate soundimages for audio alerting and feedback in the Apache Longbow (AH-64D) helicopter – Final Report*, under contract to US Army and NASA Ames Research Center, Psycho-Linguistic Research Associates, Woodside, CA, 1998.
- [4] D. Gardner and C. Simpson, *Sound image development and testing for RAH-66 (Comanche) warning caution and advisory system – Final Report*, under contract to Sikorsky Aircraft Division, Psycho-Linguistic Research Associates, Woodside, CA, 1992.
- [5] C. Stevens, D. Brennan, and S. Parker, "Simultaneous manipulation of parameters of auditory icons to convey direction, size, and distance: Effects on recognition and interpretation," in *Proc. of ICAD 04*, Sydney, Australia, July 6-9, 2004

- [6] C. Simpson and K. Marchionda-Frost, "Synthesized speech rate and pitch effects on intelligibility of warning messages for pilots, *Human Factors*, vol 26, no 5, October 1984, pp. 509-517.